

## INSTITUTE OF MATHEMATICAL ECONOMICS

Working Paper No. 363, March 2005

## THE SHAPLEY VALUE OF PHYLOGENETIC TREES

CLAUS-JOCHEN HAAKE<sup>1</sup>, AKEMI KASHIWADA<sup>2,\*</sup>, AND FRANCIS EDWARD SU<sup>2,\*\*</sup>

ABSTRACT. Every weighted tree corresponds naturally to a cooperative game that we call a *tree game*; it assigns to each subset of leaves the sum of the weights of the minimal subtree spanned by those leaves. In the context of phylogenetic trees, the leaves are species and this assignment captures the *diversity* present in the coalition of species considered. We consider the Shapley value of tree games and suggest a biological interpretation. We determine the linear transformation  $\mathbf{M}$  that shows the dependence of the Shapley value on the edge weights of the tree, and we also compute a null space basis of  $\mathbf{M}$ . Finally, we characterize the Shapley value on tree games by five axioms, a counterpart to Shapley's original theorem on the larger class of cooperative games. We also include a brief discussion of the core of tree games.

## 1. INTRODUCTION

The *Shapley value* is arguably the most important solution concept for  $n$ -player cooperative games. Given a set of players  $N$  in a cooperative game  $v$ , the Shapley value  $\varphi(N, v)$  is the unique imputation vector that satisfies four “fairness” criteria (the *Shapley axioms*) that we shall discuss later. In this paper we consider the game  $v_{\mathcal{T}}$  induced by an unrooted  $n$ -leaf tree  $\mathcal{T}$  in which each edge is assigned a positive number called an *edge weight*. In this context, the players are represented by the leaves of the tree and the value of any coalition  $S$  is the total weight of the subtree spanned by the members of  $S$ .

In a more applied context, we consider games induced by a *phylogenetic tree* in which players are species and the tree represents a proposed evolutionary relationship among the species. We suggest that a biological interpretation for the Shapley value is a notion of the average marginal diversity that a species brings to any group, and we study how the Shapley value depends on the edge weights and topology of the tree.

One possible application of the Shapley value of a phylogenetic tree is the economic theory of biodiversity preservation [4, 8]. The *Noah's ark problem* [9] asks how to prioritize species in a population if only some limited number can be saved; we suggest that Shapley value provides a natural ranking criterion.

The literature applying game-theoretic solution concepts to an analysis of trees appears to be limited. One closely related example is Kar [3], who studies cost-sharing in a network structure and characterizes the Shapley value of the minimum cost spanning tree game of an arbitrary graph. However, his work differs from ours because he considers each node of a graph as a player in the

<sup>1</sup>Institute of Mathematical Economics, Bielefeld University, PO Box 10 01 31, 33501 Bielefeld, Germany, [chaake@wiwi.uni-bielefeld.de](mailto:chaake@wiwi.uni-bielefeld.de).

<sup>2</sup>Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, U.S.A., [akashiwada@hmc.edu](mailto:akashiwada@hmc.edu), [su@math.hmc.edu](mailto:su@math.hmc.edu).

\*Research partially supported by a Howard Hughes Medical Institute Undergraduate Science Education Program grant to Harvey Mudd College.

\*\*Research partially supported by NSF Grant DMS-0301129.

game, whereas we specifically study tree games and allow only leaves as players. Day and McMorris [1] propose suitable axioms for a consensus rule that will aggregate several phylogenetic trees into one consensus tree; this differs from the thrust of our work, which is to consider one tree and explore the interpretation and properties of the Shapley value of the associated tree game.

In the next section we provide a biological interpretation for the Shapley value of phylogenetic trees. Then we discuss the mathematics of calculating the Shapley value on tree games, starting with some examples on small trees. In the subsequent section we present several theorems demonstrating how the Shapley value of an  $n$ -player game can be calculated from its  $n - m$  player subgames. We also examine the null space of the Shapley value with respect to the tree topology. In Section 6, we take a brief look at the core of tree games. We conclude this paper by developing an analogue of Shapley's theorem that characterizes the Shapley value on games by four axioms. We show that on the smaller class of tree games, the Shapley value is characterized by those four axioms plus an additional axiom.

## 2. PHYLOGENETIC TREES AND THE SHAPLEY VALUE

**2.1. Phylogenetic trees.** Evolutionary relationships between species are frequently represented by a *phylogenetic tree*. Evidence for such relationships can come from a variety of sources, such as genomic data or morphological comparisons, and much work has been done to develop methods for constructing a phylogenetic tree from such data (for surveys, see Felsenstein [2] and Semple-Steel [6]).

Phylogenetic trees are usually binary trees in which each internal node represents a bifurcation in some characteristic and the leaves are the species for which we have data. Each edge has a weight that represents some unit of distance between the nodes at its endpoints (for instance, it could be the time between speciation events). Figure 1 gives a small example of what a (rooted) phylogenetic tree could look like. However, in this paper we shall not be concerned with the location of the root of a tree, so all our trees will be unrooted.

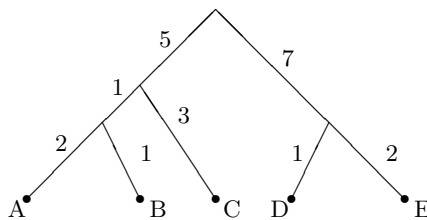


FIGURE 1. Example of a phylogenetic tree with species A-E with edge weights labeled.

Formally, we shall think of a phylogenetic tree  $\mathcal{T}$  as an unrooted tree with leaf set  $N := \{1, \dots, n\}$  (representing the species in the population), edge set  $E$ , and an edge weight  $\alpha_k$  for each edge  $k$  in  $E$ .

**2.2. The Shapley value.** In cooperative game theory, a *cooperative game* is a pair  $(N, v)$  consisting of a set of *players*  $N = \{1, 2, \dots, n\}$  and a *characteristic function*  $v$  that takes every subset of  $N$  (called a *coalition*) to a real number (called the *worth* of the coalition). The subset consisting of all players is called the *grand coalition*. Formally, if  $2^N$  is the set of all subsets of  $N$ , then  $v : 2^N \rightarrow \mathbb{R}$ .

For instance,  $N$  could be a set of companies and  $v$  could describe the profit that each coalition of companies could make if the members of that coalition worked together. Usually, game theorists are interested in *superadditive* games in which  $v(S \cup T) \geq v(S) + v(T)$  for any two coalitions  $S$  and  $T$ . In such games, there is incentive for players to cooperate when the inequality is strict. One of the basic questions in cooperative game theory is: if players work together to achieve some total worth (in our example, profit), how should players then distribute their worth (profit) among themselves?

As all (Pareto efficient) solution concepts from cooperative game theory do, the *value* introduced by Shapley [7] suggests a “fair” distribution of the total worth of the entire set of players  $N$  among the members of  $N$ . Given a cooperative game  $(N, v)$ , the Shapley value is a vector  $\varphi = (\varphi_i)$  defined by the formula

$$(1) \quad \varphi_i(N, v) = \frac{1}{n!} \sum_{\substack{S \subseteq N \\ i \in S}} (s-1)!(n-s)!(v(S) - v(S-i))$$

where  $s = |S|$  is the size of the coalition  $S$  and  $n = |N|$  is the total number of players.

The formula above has a sensible interpretation that suggests a rationale for the Shapley value to obtain a “fair” distribution. For a player  $i \in N$  and a coalition  $S \subseteq N$  that contains  $i$ , the quantity  $v(S) - v(S-i)$  describes  $i$ ’s marginal contribution to the worth of  $S$ . If we choose a random ordering of the players, and consider the growing coalition that forms when the players are added one-by-one from that ordering, then the combinatorial form of (1) reflects the Shapley value’s interpretation as an *average of the marginal contributions* that  $i$  makes to a randomly chosen coalition.

**2.3. The Phylogenetic Tree Game.** Given a phylogenetic tree  $\mathcal{T}$ , we can define an associated cooperative game  $(N, v_{\mathcal{T}})$  that we call a *phylogenetic tree game*. Let  $N$  be the set of leaves of the tree (species). For any subset  $S \subseteq N$  of species, consider the unique spanning subtree containing the members in  $S$ , and let  $v_{\mathcal{T}}(S)$  be the sum of the edge weights of that spanning tree. Thus for each set  $S$  we may think of  $v_{\mathcal{T}}(S)$  a measure of the *diversity* within  $S$ . This forms a cooperative game  $(N, v_{\mathcal{T}})$  in a natural way, and it is evident from our definition that the phylogenetic tree game is superadditive.

Although species can hardly be compared with rationally acting agents (as usually assumed in theory of cooperative games), we may still ask for a meaningful re-interpretation of game-theoretic solution concepts such as the Shapley value in the context of phylogenetic trees.

Given a phylogenetic tree game  $(N, v_{\mathcal{T}})$ , equation (1) suggests that the Shapley value of a given species may be thought of as its *average marginal diversity*, i.e., the average diversity the species can be expected to add to a group that it joins. So if  $\varphi_i > \varphi_j$ , then species  $i$  can be thought to contribute a greater diversity to a group than species  $j$  might.

**2.4. The Shapley Value Axioms.** Besides the interpretation of the Shapley value as an average expected marginal contribution, there is an axiomatization of the Shapley value (see [7]) that uniquely characterizes it by a set of (desirable) properties. We review the axioms presented by Shapley and discuss their plausibility in the present setting as properties of phylogenetic trees. Let therefore  $\mathcal{V} := \{v : 2^N \rightarrow \mathbb{R} \mid v(\emptyset) = 0\}$  be the set of all cooperative games with  $n$  players.

- (1) (*Pareto Efficiency Axiom*) The Shapley value is Pareto efficient, i.e.,  $\sum_{i \in N} \varphi_i(N, v) = v(N)$  for all  $v \in \mathcal{V}$ .

This axiom just states that the total diversity present within a phylogenetic tree will be distributed and ascribed to the species within it. This is a reasonable axiom, given that

the purpose of a solution concept for a cooperative game is to distribute the worth of the grand coalition among its members. In this context, the natural interpretation is that the Shapley value answers the question of how much a specific species is responsible for the total diversity, or, put another way, what is its *share* of  $v_{\mathcal{T}}(N)$ .

- (2) (*Symmetry Axiom*) For any permutation of players  $\pi : N \rightarrow N$  the Shapley value satisfies  $\varphi(\pi v) = \pi \varphi(v)$ , where  $\pi v$  is the permuted game given by  $\pi v(S) := v(\pi^{-1}(S))$  for all  $S \subseteq N$  and  $\pi \varphi(v)$  is the permuted solution vector, i.e.,  $(\pi \varphi(v))_i := \varphi_{\pi^{-1}(i)}(v)$ .

The symmetry axiom states that a player's allocation should not be based on her name. Another consequence of the symmetry axiom is if exchanging two players causes no difference in the worth that each adds to any coalition, then they should have the same Shapley value. Biologically speaking, if two species play the same role within a tree then they should be ascribed the same responsibility for diversity, which seems to be a plausible requirement.

- (3) (*Dummy Axiom*) A dummy player is one that does not add worth to the value of any coalition. This axiom says that dummy players should have a Shapley value of zero.

This axiom is vacuously satisfied in the case of a phylogenetic tree game because there are no dummy species. To see this, note that every species  $i$  adds worth to the coalition that consists of a single species  $j \neq i$ , because the weight of the subtree containing  $i$  and  $j$  is the sum of the edge weights between  $i$  and  $j$  and is therefore non-zero, but the weight of the subtree consisting of the singleton  $j$  is zero. (Even though there are no dummy species, this is still a reasonable axiom here, since any species that does not diversify any coalition should get value zero.)<sup>1</sup>

- (4) (*Additivity Axiom*) Given two games  $(N, v)$  and  $(N, w)$  in  $\mathcal{V}$  with the same set of players  $N$ , define the *sum game*  $(N, v + w)$  with characteristic function  $(v + w)(S) = v(S) + w(S)$  for every coalition  $S$ . This axiom stipulates that the Shapley value of the sum game should be the sum of the Shapley values of the individual games:  $\varphi(N, v + w) = \varphi(N, v) + \varphi(N, w)$ .

As an example, suppose we are given genome sequences for a set of species  $N$ , and each sequence has length 200. For each pair of species  $i, j$  consider the (rather crude) measure of distance  $d(i, j)$  to be the number of positions in which the sequences differ. The pairwise distance data can be used to construct a tree (using any standard method) and consequently, a tree game. Thus the first 100 positions of the sequences can be used to construct a tree game  $(N, v_1)$ , and the second 100 positions a tree game  $(N, v_2)$ . Then the Shapley value of the sum game  $(N, v_1 + v_2)$  is the sum of the Shapley values for each game. This seems plausible in this context, since if the pairwise distances  $d(i, j)$  from both sets of 100 positions actually arise from a tree metrics on the same tree, then the sum game will arise from the tree reconstructed from all 200 positions.

### 3. EXAMPLES AND MOTIVATION: THE SHAPLEY VALUE FOR SMALL TREES

As can be seen from (1), the Shapley value of a tree game is a linear function of the edge weights of the tree. We call that linear transformation the *Shapley transformation*. Before deriving a general formula for this transformation in the subsequent section, we study the Shapley transformation for games induced by unrooted three-, four-, five- and six-leaf trees.

---

<sup>1</sup>In Section 5 we will replace the dummy axiom by a different one to characterize the Shapley value on the class of games that actually come from trees.

We will refer to the weights of edges incident to leaves as *leaf weights* and other edge weights as *internal edge weights*. Note that for an unrooted  $n$ -leaf tree, there are  $n - 2$  internal nodes and  $n - 3$  internal edges in  $E$ . In what follows, the superscript  $T$  denotes the *transpose*.

**Definition 3.1.** Let  $\mathcal{T}$  be an  $n$ -leaf tree with leaves  $N = \{1, \dots, n\}$ , associated leaf weights  $\alpha_1, \dots, \alpha_n$  and internal edges  $I_1, \dots, I_{n-3}$  with associated internal edge weights  $\alpha_{I_1}, \dots, \alpha_{I_{n-3}}$ . Let  $\vec{E}$  be a vector consisting of the edge weights in this order:  $(\alpha_1, \dots, \alpha_n, \alpha_{I_1}, \dots, \alpha_{I_{n-3}})^T$ . Define  $\mathbf{M} = \mathbf{M}(N, v_{\mathcal{T}})$  to be the  $n \times (2n - 3)$  matrix that represents the Shapley transformation, so that the Shapley value of the game  $v_{\mathcal{T}}$  is

$$\varphi(N, v_{\mathcal{T}}) = (\varphi_1, \varphi_2, \dots, \varphi_n)^T = \mathbf{M}\vec{E}$$

where  $\varphi_i$  is the Shapley value associated with leaf  $i$ . Note that  $\mathbf{M}$  depends on the topology of the  $n$ -leaf tree.

Later we will determine a formula for  $\mathbf{M}[i, k]$ , which is the coefficient of edge weight  $k$  in the calculation of the Shapley value of  $i$ . But first, we give a few examples.

**3.1. Three-Leaf Trees.** Topologically, there is only one unrooted three-leaf tree  $\mathcal{T}$ . Let the leaves represent players A, B, and C with corresponding leaf weights  $\alpha$ ,  $\beta$ , and  $\gamma$  as seen in Figure 2.

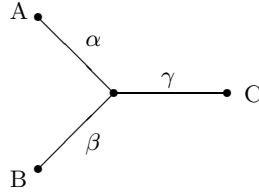


FIGURE 2. The topology of an unrooted three-leaf tree  $\mathcal{T}$  where the players are A, B, and C with corresponding leaf weights  $\alpha$ ,  $\beta$ , and  $\gamma$ .

The characteristic function  $v_{\mathcal{T}}$  for this game is

$$\begin{aligned} v_{\mathcal{T}}(A) &= v_{\mathcal{T}}(B) = v_{\mathcal{T}}(C) = 0, \\ v_{\mathcal{T}}(AB) &= \alpha + \beta, \quad v_{\mathcal{T}}(AC) = \alpha + \gamma, \quad v_{\mathcal{T}}(BC) = \beta + \gamma, \\ v_{\mathcal{T}}(ABC) &= \alpha + \beta + \gamma. \end{aligned}$$

Using Definition 3.1, we can calculate the Shapley value by  $\varphi = (\varphi_A, \varphi_B, \varphi_C) = \mathbf{M}\vec{\ell}$  where  $\vec{\ell}$  is the vector of leaf weights  $(\alpha, \beta, \gamma)^T$  and

$$\mathbf{M} = \frac{1}{6} \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix}.$$

It is apparent that we can solve for  $\alpha$ ,  $\beta$ , and  $\gamma$  in terms of  $\varphi$  by inverting  $\mathbf{M}$ :

$$\vec{\ell} = \frac{1}{3} \begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix} \begin{pmatrix} \varphi_A \\ \varphi_B \\ \varphi_C \end{pmatrix}.$$

This means the Shapley value of a 3-leaf tree uniquely determines the tree representing the game.

**3.2. Four- and Five-Leaf Trees.** Using the same procedure as in the three-leaf tree case, we can calculate the Shapley value for each player in the four- and five-leaf case. There is a unique tree topology for each case as shown in figure 3.

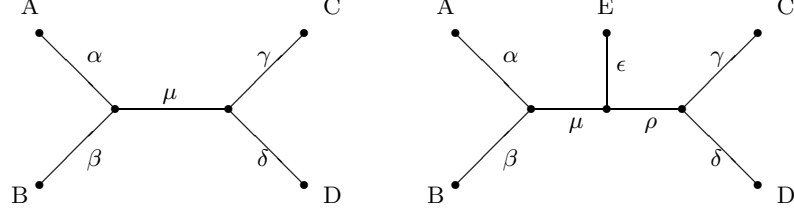


FIGURE 3. (*left*) The topology for an unrooted four-leaf tree where the players are A, B, C, and D. (*right*) The unrooted five-leaf tree with players A, B, C, D, and E.

The Shapley value for the general four-leaf tree game is

$$\frac{1}{24} \begin{bmatrix} 18 & 2 & 2 & 2 & 6 \\ 2 & 18 & 2 & 2 & 6 \\ 2 & 2 & 18 & 2 & 6 \\ 2 & 2 & 2 & 18 & 6 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \mu \end{pmatrix}.$$

Similarly for the five-leaf tree game, the Shapley value is

$$\frac{1}{120} \begin{bmatrix} 96 & 6 & 6 & 6 & 6 & 36 & 16 \\ 6 & 96 & 6 & 6 & 6 & 36 & 16 \\ 6 & 6 & 96 & 6 & 6 & 16 & 36 \\ 6 & 6 & 6 & 96 & 6 & 16 & 36 \\ 6 & 6 & 6 & 6 & 96 & 16 & 16 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \mu \\ \rho \end{pmatrix}.$$

It is apparent from the fact that there are more variables (edge weights) than equations that there is not a unique set of (possibly negative) edge weights for a given Shapley value. That is, there is not a unique tree corresponding to a given Shapley value. The null space of  $\mathbf{M}$  will therefore help us determine which weighted trees have the same Shapley value. A basis for the null space of  $\mathbf{M}$  for the four-leaf tree is

$$\left\{ \begin{pmatrix} -1/4 \\ -1/4 \\ -1/4 \\ -1/4 \\ 1 \end{pmatrix} \right\}$$

This means that given a tree  $\mathcal{T}$ , we can produce other trees with the same Shapley value by reducing the leaf weights by  $1/4$  for each unit increase in the internal edge weight.

Similarly, a null space basis for the five-leaf tree is

$$\left\{ \begin{pmatrix} -1/3 \\ -1/3 \\ -1/9 \\ -1/9 \\ -1/9 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/9 \\ -1/9 \\ -1/3 \\ -1/3 \\ -1/3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

**3.3. Six-Leaf Trees.** For our last direct calculation, let us consider the games represented by six-leaf trees. In this case there are two topologies for unrooted trees with six leaves (see figure 4).

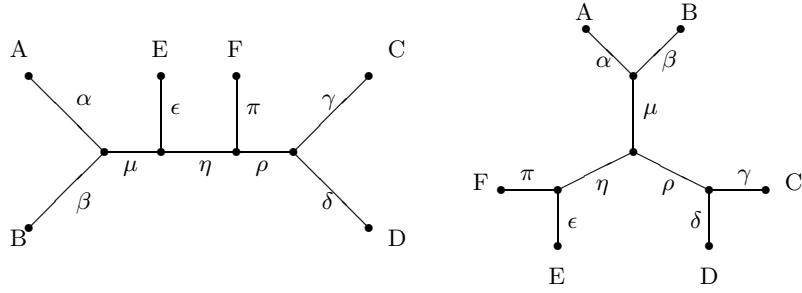


FIGURE 4. (*left*) The first topology for an unrooted six-leaf tree  $\mathcal{T}$  where the players are A, B, C, D, E and F. (*right*) The second unrooted six-leaf tree  $\mathcal{T}'$ .

The Shapley value for the first and second six-leaf trees are, respectively,

$$\varphi(N, v_{\mathcal{T}}) = \frac{1}{720} \begin{bmatrix} 600 & 24 & 24 & 24 & 24 & 24 & 240 & 60 & 120 \\ 24 & 600 & 24 & 24 & 24 & 24 & 240 & 60 & 120 \\ 24 & 24 & 600 & 24 & 24 & 24 & 60 & 240 & 120 \\ 24 & 24 & 24 & 600 & 24 & 24 & 60 & 240 & 120 \\ 24 & 24 & 24 & 24 & 600 & 24 & 60 & 60 & 120 \\ 24 & 24 & 24 & 24 & 24 & 600 & 60 & 60 & 120 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \pi \\ \mu \\ \rho \\ \eta \end{pmatrix},$$

$$\varphi(N, v_{\mathcal{T}'}) = \frac{1}{720} \begin{bmatrix} 600 & 24 & 24 & 24 & 24 & 24 & 240 & 60 & 60 \\ 24 & 600 & 24 & 24 & 24 & 24 & 240 & 60 & 60 \\ 24 & 24 & 600 & 24 & 24 & 24 & 60 & 240 & 60 \\ 24 & 24 & 24 & 600 & 24 & 24 & 60 & 240 & 60 \\ 24 & 24 & 24 & 24 & 600 & 24 & 60 & 60 & 240 \\ 24 & 24 & 24 & 24 & 24 & 600 & 60 & 60 & 240 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \pi \\ \mu \\ \rho \\ \eta \end{pmatrix}.$$

As with the four and five leaf cases, both topologies of the six leaf tree allow for many trees to possess the same Shapley value. The basis for the null space of the first six-leaf tree is

$$\left\{ \begin{pmatrix} -3/8 \\ -3/8 \\ -1/16 \\ -1/16 \\ -1/16 \\ -1/16 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/16 \\ -1/16 \\ -3/8 \\ -3/8 \\ -1/16 \\ -1/16 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/6 \\ -1/6 \\ -1/6 \\ -1/6 \\ -1/6 \\ -1/6 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

and for the second six-leaf tree is

$$\left\{ \begin{pmatrix} -3/8 \\ -3/8 \\ -1/16 \\ -1/16 \\ -1/16 \\ -1/16 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/16 \\ -1/16 \\ -3/8 \\ -3/8 \\ -1/16 \\ -1/16 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/16 \\ -1/16 \\ -1/16 \\ -1/16 \\ -3/8 \\ -3/8 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

**3.4. Notes on Relationship between Trees and Shapley Values.** From these examples, we make a few observations.

- (1) Any Shapley value  $n$ -vector can be realized by adjusting the edge weights of an  $n$ -leaf tree. This may involve positive as well as nonpositive edge weights. However, the positive hull of the column vectors of the matrix  $\mathbf{M}$  can be realized as the Shapley value of some tree with nonnegative edge weights.
- (2) When  $n \geq 4$ , there is not a unique  $n$ -leaf tree corresponding to a given Shapley value because the null space is nontrivial.
- (3) The null space bases for the two six-leaf trees are different; hence enough to determine the topology of the tree. As we shall see in the next section, this phenomenon is true in general.
- (4) Under close inspection, one notices a relationship between the numbers of leaves on each side of an internal edge and quantities such as the entries of the Shapley transformation matrix and the null space basis vectors. We exhibit their explicit dependence in the next section.

#### 4. CALCULATING THE SHAPLEY VALUE FROM SUBTREES

In this section, we shall prove that the Shapley value for an  $n$ -leaf tree game can be calculated from the Shapley value of all its  $(n - m)$ -leaf subtree games. First we will show that given a tree game with  $n$  players, the Shapley values can be calculated from the Shapley value for all  $(n - 1)$ -leaf subtrees. This kind of “reconstruction” result stands in stark contrast to a result of Pachter-Speyer [5] for trees; they show that an  $n$ -leaf tree cannot necessarily be reconstructed from the weights of its  $(n - 1)$ -leaf subtrees.



We first show the contribution of each edge weight to the Shapley value; these are the entries of the matrix representing the Shapley transformation.

**4.1. Entries in Shapley Value Matrix.** The following theorem gives us a quick way of finding the  $(i, k)$ th entry of the Shapley value matrix of an  $n$ -leaf tree game. Before we state and prove the theorem, we need to present a definition that is instrumental throughout the rest of this paper.

**Definition 4.1.** Let  $\mathcal{T}$  be an  $n$ -leaf tree with leaves  $N$  and edges  $E$ . For  $i \in N$  and  $k \in E$ , the removal of edge  $k$  splits  $\mathcal{T}$  into two subtrees. Let  $\mathcal{C}(i, k)$  denote the subtree that contains  $i$  (the “containing” subtree) and let  $\mathcal{F}(i, k)$  denote the subtree that is “far” from  $i$ . We then denote the number of leaves of  $\mathcal{C}(i, k)$  and  $\mathcal{F}(i, k)$  as  $c(i, k)$  and  $f(i, k)$ , respectfully.

If it is obvious what leaf  $i$  and edge  $k$  we are referring to, we will simply write  $c, f$  instead of  $c(i, k), f(i, k)$ . Note that  $n = c + f$ . We call  $c, f$  the *split counts* associated with leaf  $i$  and edge  $k$ . As we shall see, the split counts will arise frequently in our results on the Shapley transformation.

**Theorem 4.1.** Let  $\mathcal{T}$  be an  $n$ -leaf tree. The  $(i, k)$ th entry of the Shapley transformation matrix  $\mathbf{M}$  is given by

$$\mathbf{M}[i, k] = \frac{f(i, k)}{n c(i, k)}.$$

*Proof.* It is sufficient to show this theorem is true in calculating the Shapley value of a single leaf in the  $n$ -leaf tree game. Fix leaf  $i$ . To count the number of times a given edge weight contributes to  $i$ ’s Shapley value, we need to know how many times it is in the marginal contribution of  $i$  for coalitions of size  $s$ . Edge weight  $\alpha_k$  will be part of  $i$ ’s marginal contribution if the other  $s - 1$  members of the coalition are from the opposite side of the edge from  $i$ . So

$$\mathbf{M}[i, k] = \frac{1}{n!} \sum_{s=2}^n (s-1)!(n-s)! \binom{f(i, k)}{s-1} = \frac{1}{n!} \sum_{s=2}^n \frac{(n-s)! f(i, k)!}{(f(i, k) - s + 1)!}.$$

Using the fact  $f = n - c$ , the above expression can be rewritten:

$$\frac{1}{n!} \sum_{s=2}^n (n-c)!(c-1)! \binom{n-s}{c-1} = \frac{(n-c)!(c-1)!}{n!} \sum_{j=1}^{n-1} \binom{j-1}{c-1}.$$

We use the identity

$$\sum_{j=1}^n \binom{j-1}{c-1} = \binom{n}{c} = \binom{n-1}{c-1} \frac{f}{c} + \binom{n-1}{c-1}$$

to obtain

$$\mathbf{M}[i, k] = \frac{(n-c)!(c-1)!}{n!} \binom{n-1}{c-1} \frac{f}{c} = \frac{f}{nc}.$$

□

This result is particularly nice because it shows how the Shapley value’s dependence on any edge weight only depends on the number of leaves on either side of that edge. Consider the following example.

**Example 4.1.** Using Theorem 4.1 we will calculate the coefficient of  $\mu$  in player  $A$ ’s Shapley value for a five-leaf tree. Let the edge with edge weight  $\mu$  be  $I_1$ . There are three leaves in  $\mathcal{F}(A, I_1)$  and two leaves in  $\mathcal{C}(A, I_1)$ . Thus,

$$\mathbf{M}[1, 6] = \frac{3}{5 \cdot 2}$$

which is the same as the  $(A, \mu)$  entry 36/120 in the Shapley transformation of the five-leaf tree given in section 3.2.

With the above result we can calculate the Shapley value of an  $n$ -leaf tree game from the Shapley value of subtree games.

**4.2. Shapley Value from  $(n-1)$ -leaf Subtrees.** In this section we want to show how the Shapley value of an  $n$ -leaf tree game can be calculated from the Shapley value of its  $(n-1)$ -leaf subtrees. Before we can do that we need the following definition and lemma.

**Definition 4.2.** Let  $\mathcal{T}$  be an  $n$ -leaf tree with leaves  $N$ . For any subset of leaves  $S \subseteq N$ , the Shapley value of  $N$  with respect to the subtree spanned by  $S$  is  $\varphi(N, v_{S, \mathcal{T}}) \in \mathbb{R}^n$  where for any coalition  $U \subseteq N$ ,  $v_{S, \mathcal{T}}(U) = v_{\mathcal{T}}(S \cap U)$ . Put another way,  $\varphi(N, v_{S, \mathcal{T}}) = \varphi(S, v_{S, \mathcal{T}}) \times \{0\}^{N \setminus S}$ .

In other words, this Shapley value will assign zero to any player not in  $S$  and the usual Shapley value of the tree game spanned by  $S$  to players in  $S$ .

**Lemma 4.1.** Let  $\mathcal{T}$  be an  $n$ -leaf tree with leaves  $N = \{1, \dots, n\}$  and internal edges  $I_1, \dots, I_{n-3}$  with corresponding edge weights  $\alpha_{I_1}, \dots, \alpha_{I_{n-3}}$ . Fix  $i \in N$ . Then

$$\sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \sum_{k=1}^{n-3} \frac{f_S(i, k)}{c_S(i, k)} \alpha_{I_k} = (n-1) \sum_{k=1}^{n-3} \frac{f_N(i, k)}{c_N(i, k)} \alpha_{I_k}$$

where  $f_U, c_U$  are determined with respect to leaf  $i$  in the tree spanned by  $U \subseteq N$ .

*Proof.* Fix  $k \in \{1, \dots, n-3\}$ . Notice that each subset  $S$  is obtained by deleting one leaf  $j \neq i$  either from  $\mathcal{F}(i, I_k)$  or from  $\mathcal{C}(i, I_k)$ . Since  $|\mathcal{F}(i, I_k)| = f_N$  and  $|\mathcal{C}(i, I_k)| = c_N - 1$ ,

$$\sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \frac{f_S}{c_S} \alpha_{I_k} = \left( f_N \frac{f_N - 1}{c_N} + (c_N - 1) \frac{f_N}{c_N - 1} \right) \alpha_{I_k} = (n-1) \frac{f_N}{c_N} \alpha_{I_k}.$$

Thus, summing over all  $k$ , we obtain

$$\sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \sum_{k=1}^{n-3} \frac{f_S}{c_S} \alpha_{I_k} = (n-1) \sum_{k=1}^{n-3} \frac{f_N}{c_N} \alpha_{I_k}.$$

□

Now we are ready to show how we can calculate the Shapley value of an  $n$ -leaf tree from the Shapley values for all its  $(n-1)$ -leaf subtrees.

**Theorem 4.2.** Let  $\mathcal{T}$  be an unrooted  $n$ -leaf tree with leaves  $N = \{1, \dots, n\}$  and corresponding leaf weights  $\alpha_1, \dots, \alpha_n$ . Similarly, label the internal edges  $I_1, \dots, I_{n-3}$  with edge weights  $\alpha_{I_1}, \dots, \alpha_{I_{n-3}}$ . If the Shapley values for all  $(n-1)$ -leaf subtrees are known, then the Shapley value for  $N$  is

$$(2) \quad \varphi(N, v_{\mathcal{T}}) = \frac{1}{n} \left( \vec{\ell} + \sum_{\substack{S \subseteq N \\ |S|=n-1}} \varphi(N, v_{S, \mathcal{T}}) \right),$$

where  $\vec{\ell}$  is the vector of leaf weights  $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ .

*Proof.* First we will show this theorem is true when calculating the Shapley value for one leaf. Fix  $i \in \{1, \dots, n\}$ . Using Theorem 4.1,

$$\begin{aligned} \varphi_i(N, v_T) &= \frac{1}{n} \left( (n-1)\alpha_i + \sum_{\substack{j \in \{1, \dots, n\} \\ i \neq j}} \frac{1}{n-1} \alpha_j + \sum_{k=1}^{n-3} \frac{f_N(i, k)}{c_N(i, k)} \alpha_{I_k} \right) \\ (3) \quad &= \frac{1}{n} \left( \alpha_i + \frac{1}{n-1} \left( (n-1)(n-2)\alpha_i + \sum_{\substack{j \in \{1, \dots, n\} \\ i \neq j}} \frac{n-2}{n-2} \alpha_j + (n-1) \sum_{k=1}^{n-3} \frac{f_N(i, k)}{c_N(i, k)} \alpha_{I_k} \right) \right) \end{aligned}$$

where  $f_N, c_N$  are determined with respect to  $\mathcal{T}$ .

Let  $T'$  be the subtree of  $\mathcal{T}$  spanned by  $N \setminus \{x\}$  where  $x \neq i$ . Again, using Theorem 4.1,

$$\varphi_i(N \setminus \{x\}, v_{T'}) = \frac{1}{n-1} \left( (n-2)\alpha_i + \sum_{\substack{j \in \{1, \dots, n\} \\ i \neq j \neq x}} \frac{1}{n-2} \alpha_j + \sum_{k=1}^{n-3} \frac{f_{N \setminus \{x\}}(i, k)}{c_{N \setminus \{x\}}(i, k)} \alpha_{I_k} \right)$$

where  $f_{N \setminus \{x\}}, c_{N \setminus \{x\}}$  are determined with respect to  $T'$ .

We can see that  $i$  is a member of  $n-1$  of the  $(n-1)$ -leaf subtrees and every other leaf is in  $n-2$  of those subtrees. Using these facts and Lemma 4.1, we can rewrite (3) as

$$\frac{1}{n} \left( \alpha_i + \sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \varphi_i(N, v_S) \right).$$

Therefore,

$$\varphi(N, v_T) = \frac{1}{n} \left( \vec{\ell} + \sum_{\substack{S \subseteq N \\ |S|=n-1}} \varphi(N, v_S) \right).$$

□

**Example 4.2.** Consider the five-leaf tree  $\mathcal{T}$  in figure 5. From direct calculations we see that

$$\varphi(\{A, B, C, D, E\}, v_T) = (5.28, 6.78, 4.2, 4.95, 2.78).$$

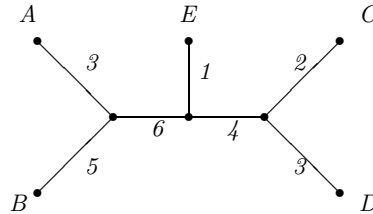


FIGURE 5. Example of calculating the Shapley value of a five-leaf tree from all 4-leaf subtrees.

The Shapley value for each four-leaf subtree is

$$\begin{aligned}\varphi(N, v_{\{A,B,C,D\}}) &= (5.58, 6.92, 4.92, 5.58, 0), \\ \varphi(N, v_{\{A,B,C,E\}}) &= (4.75, 6.08, 6.75, 0, 3.42), \\ \varphi(N, v_{\{A,B,D,E\}}) &= (4.83, 6.17, 0, 7.5, 3.5), \\ \varphi(N, v_{\{A,C,D,E\}}) &= (8.25, 0, 3.58, 4, 25, 2.92), \\ \varphi(N, v_{\{B,C,D,E\}}) &= (0, 9.75, 3.75, 4.42, 3.08).\end{aligned}$$

Using (2), we get

$$\begin{aligned}\varphi(\{A, B, C, D, E\}, v_T) &= \frac{1}{5} \left( \begin{pmatrix} 3 \\ 5 \\ 2 \\ 3 \\ 1 \end{pmatrix} + \sum_{\substack{S \subseteq N \\ |S|=n-1}} \varphi(N, v_S) \right) \\ &= \frac{1}{5} \left( \begin{pmatrix} 3 \\ 5 \\ 2 \\ 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 23.41 \\ 28.92 \\ 19 \\ 21.75 \\ 12.92 \end{pmatrix} \right) \\ &= (5.28, 6.78, 4.2, 4.95, 2.78).\end{aligned}$$

**4.3. Generalizing Theorem 4.2.** Now that we have looked at calculating the Shapley value from the  $(n-1)$ -leaf subtrees of a game tree, it would be nice to generalize the formula so we can use any size subtrees. Although it looks as if it would be easy to induct on (2), it is a bit tricky when it comes to figuring out what the entries of  $\vec{\ell}$  should be. In some cases, the  $i$ th entry of  $\vec{\ell}$  will be a sum of internal edge weights with  $i$ 's leaf weight. The following example illustrates this situation.

**Example 4.3.** In the case of a five-leaf tree, we can calculate the Shapley value for  $A$  from (2) by

$$\varphi_A(N, v_T) = \frac{1}{5} \left( \alpha + \sum_{\substack{S \subseteq N \\ |S|=4}} \varphi_A(N, v_S) \right).$$

If we want to calculate the Shapley value for  $A$  from the three-leaf subtrees we obtain

$$\begin{aligned}\varphi_A(N, v_T) &= \frac{1}{5} \left( \alpha + \sum_{\substack{S \subseteq N \\ |S|=4}} \frac{1}{4} (\alpha' + \sum_{\substack{U \subseteq S \\ |U|=3}} \varphi_A(N, v_U)) \right) \\ &= \frac{1}{5} \left( \alpha + \frac{1}{4} \left( 4\alpha + \mu + 2 \sum_{\substack{U \subseteq S \\ |U|=3}} \varphi_A(N, v_U) \right) \right) \\ &= \frac{1}{5} \left( 2\alpha + \frac{1}{4} \left( \mu + 2 \sum_{\substack{U \subseteq S \\ |U|=3}} \varphi_A(N, v_U) \right) \right).\end{aligned}$$

The summand  $\frac{1}{4}\mu$  came from the factor of  $\vec{\ell}$  from the subtree ACDE. See figure 6. In this case, the leaf weight  $\alpha'$  of A is  $\alpha + \mu$ .

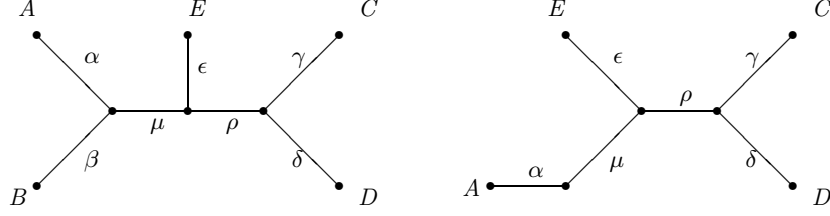


FIGURE 6. (left) The five-leaf tree where the players are A, B, C, D, and E. (right) The four-leaf subtree ACDE. Notice that the leaf weight of A is now  $\alpha + \mu$  instead of just  $\alpha$ .

Taking the internal edge weights into account, we get an equation for the Shapley value from  $(n - m)$ -leaf subtrees.

**Theorem 4.3.** *Let  $N = \{1, \dots, n\}$  be the leaves of tree  $\mathcal{T}$  and label the internal edges  $\{n+1, \dots, 2n-3\}$ . Let the associated edge weights be  $\alpha_k$  for  $k \in \{1, \dots, 2n-3\}$ . If all of the Shapley values for the  $(n - m)$ -leaf subtrees are known, then the Shapley value of  $\mathcal{T}$  is*

$$(4) \quad \varphi(N, v_{\mathcal{T}}) = \frac{1}{n!} \left( \vec{L}(m) + (n - m)!m! \sum_{\substack{S \subseteq N \\ |S|=n-m}} \varphi(N, v_S) \right)$$

where

$$(5) \quad \vec{L}_i(m) = \sum_{k=1}^{2n-3} \sum_{j=c(i,k)}^m (n - j)!(j - 1)! \binom{f(i,k)}{n - j} \alpha_k.$$

*Proof.* We will prove this by induction on  $m$ . It suffices to prove this for a single leaf so fix  $i \in N$ .

Base case: When  $m = 1$ , we have

$$\begin{aligned} & \frac{1}{n!} \left( \vec{L}_i(1) + (n - 1)!1! \sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \varphi_i(N, v_S) \right) \\ &= \frac{1}{n!} \left( (n - 1)!(1 - 1)! \binom{1}{n - 1} \alpha_i + (n - 1)! \sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \varphi_i(N, v_S) \right) \\ &= \frac{1}{n} \left( \alpha_i + \sum_{\substack{i \in S \subseteq N \\ |S|=n-1}} \varphi_i(N, v_S) \right). \end{aligned}$$

By theorem 4.2 this is  $\varphi_i(N, v_{\mathcal{T}})$ .

Induction Hypothesis:

$$\varphi_i(N, v_T) = \frac{1}{n!} \left( \vec{L}_i(m-1) + (n-m+1)!(m-1)! \sum_{\substack{i \in S \subseteq N \\ |S|=n-m+1}} \varphi_i(N, v_S) \right).$$

Inductive Step: We can apply theorem 4.2 to the induction hypothesis to get

$$(6) \quad \varphi_i(N, v_T) = \frac{1}{n!} \left( \vec{L}_i(m-1) + (n-m)!(m-1)! \sum_{\substack{S \subseteq N \\ |S|=n-m+1}} \left( \alpha'_i + \sum_{\substack{i \in U \subseteq N \\ |U|=n-m}} \varphi_i(N, v_U) \right) \right)$$

where  $\alpha'_i$  is the leaf weight of  $i$  in the subtree spanned by  $U$ . Notice that any edge weight is in  $\alpha'_i$  if  $U \setminus \{i\}$  is from the opposite side of that edge from  $i$ . Thus

$$(n-m)!(m-1)! \sum_{\substack{S \subseteq N \\ |S|=n-m+1}} \alpha'_i = \sum_{k=1}^{2n-3} (n-m)!(m-1)! \binom{f(i,k)}{n-m} \alpha_k.$$

If we add this to  $\vec{L}_i(m-1)$  we get  $\vec{L}_i(m)$ . Also note that every  $(n-m)$ -leaf subtree comes from  $m$   $(n-m+1)$ -leaf trees so plugging that into (6) yields

$$\varphi_i(N, v_T) = \frac{1}{n!} \left( \vec{L}_i(m) + (n-m)!m! \sum_{\substack{i \in U \subseteq N \\ |U|=n-m}} \varphi_i(N, v_U) \right).$$

□

It is interesting to note that (4) does not seem to depend on the topology of the tree so we can theoretically induct on the size of the subtrees. However, with the addition of (5) we lose the ability to induct since we need to know the topology of the tree or the split counts. Equation (4) is helpful in seeing how the Shapley value depends on each of the edge weights. We can see this from the following corollary.

**Corollary 4.1.** *If  $N$  is the set of leaves in tree  $\mathcal{T}$  with edge weights  $\alpha_k$  for  $k = 1, \dots, 2n-3$ , then for  $i \in N$ ,*

$$\varphi_i(N, v_T) = \frac{1}{n!} \left( \sum_{k=1}^{2n-3} \sum_{j=c(i,k)}^{n-2} (n-j)!(j-1)! \binom{f(i,k)}{n-j} \alpha_k \right).$$

*Proof.* Use theorem 4.3 when  $m = n-2$  and

$$\sum_{\substack{S \subseteq N \\ |S|=2}} \varphi_i(S, v_T) = \sum_{k=1}^{2n-3} \frac{1}{2} \alpha_k f(i, k).$$

□

**4.4. Examining the Null Space.** As we have seen, Theorem 4.1 has been instrumental in showing how the Shapley value can be calculated from the Shapley value of any size subtrees. Now we will also use this theorem to understand the dependence of the null space on the split counts, as mentioned in section 3.4.

The following theorem exhibits the null spaces basis of  $\mathbf{M}$  in terms of the split counts.

**Theorem 4.4.** *Let  $\mathcal{T}$  be an  $n$ -leaf tree with leaves  $N = \{1, \dots, n\}$  and internal edges  $I_1, \dots, I_{n-3}$ . For each internal edge  $I_k$ , there corresponds a vector  $w_{I_k} \in \mathbb{R}^{2n-3}$  in a basis of the null space of the Shapley transformation of  $\mathcal{T}$ :*

$$(7) \quad (w_{I_k})_i = \begin{cases} -\frac{f(i,k)-1}{(n-2)c(i,k)} & \text{if } 1 \leq i \leq n \\ 1 & \text{if } i = n+k \\ 0 & \text{otherwise} \end{cases}$$

for all  $k \in \{1, \dots, n-3\}$  and entries  $i \in \{1, \dots, 2n-3\}$ , where the first  $n$  entries correspond to leaves and the last  $n-3$  entries corresponds to internal edges.

Before proving the theorem, we give an example.

**Example 4.4.** *Consider the five-leaf tree in Figure 3. Label the internal edges  $I_1, I_2$  such that the corresponding edge weights are  $\mu, \rho$ , respectively. Using Theorem 4.4, let us calculate the null space vector  $w_{I_1}$ . We know that the  $5+1=6$ th entry of  $w_{I_1}$  is 1 and all entries after that are zero. To find the first five entries of the vector, we consider the two subtrees obtained by removing  $I_1$  from the tree. In that case, we'll get the subtrees  $AB$  and  $CDE$ . Then using (7), the first two entries of the matrix corresponding to  $A$  and  $B$  will be*

$$-\frac{3-1}{(5-2)2} = -\frac{1}{3}$$

and the next three entries corresponding to  $C, D$ , and  $E$  are

$$-\frac{2-1}{(5-2)3} = -\frac{1}{9}.$$

These values correspond to the first vector in the null space basis we presented in Section 3.2. We may obtain the other basis vector in a similar fashion, by considering edge  $I_2$  instead of  $I_1$ .

Now we will prove Theorem 4.4.

*Proof.* Let  $\mathcal{T}$  be an  $n$ -leaf tree. Consider the  $i$ th leaf. If we let  $M$  be the matrix of Shapley value coefficients for  $\mathcal{T}$  then we want to show

$$(8) \quad \sum_{j=1}^{2n-3} \mathbf{M}[i, j] (w_{I_k})_j = 0.$$

Fix  $k \in \{1, \dots, n-3\}$ . There are a couple of notes to point out that make this proof easier. First using Theorem 4.1, for all leaves  $j \neq i$ ,

$$\mathbf{M}[i, j] = (n-2)!$$

and

$$\mathbf{M}[i, i] = (n-1)(n-1)!$$

for  $j \in \{1, \dots, n\}$  (not including the factor of  $\frac{1}{n!}$ ). The only other entry of the matrix we need to consider is  $M[i, n+k]$  since our construction of  $w_{I_k}$  has zeros for the rest of the entries. Thus

$$\mathbf{M}[i, n+k] = (n-1)! \frac{f(i, n+k)}{c(i, n+k)}.$$

Plugging all of this into (8) yields

$$\begin{aligned} \sum_{j=1}^{2n-3} \mathbf{M}[i, j](w_{I_k})_j &= -f(n-2)! \frac{c-1}{(n-2)f} - (c-1)(n-2)! \frac{f-1}{(n-2)c} \\ &\quad - (n-1)(n-1)! \frac{f-1}{(n-2)c} + (n-1)! \frac{f}{c}. \end{aligned}$$

To show this is the same as showing

$$(9) \quad (n-1)! \frac{f}{c} = f(n-2)! \frac{c-1}{(n-2)f} + (c-1)(n-2)! \frac{f-1}{(n-2)c} + (n-1)(n-1)! \frac{f-1}{(n-2)c}.$$

The right side of the equation (9) is

$$\begin{aligned} &f(n-2)! \frac{c-1}{(n-2)f} + (c-1)(n-2)! \frac{f-1}{(n-2)c} + (n-1)(n-1)! \frac{f-1}{(n-2)c} \\ &= (n-2)! \left( \frac{c-1+f-1}{n-2} \right) + (n-1)(n-1)! \frac{f-1}{(n-2)c} - (n-2)! \frac{f-1}{(n-2)c} \\ &= (n-2)! \left( \frac{(c-1)(n-1)}{(n-2)c} \right) + (n-1)(n-1)! \frac{f-1}{(n-2)c} \\ &= \frac{(n-1)!}{(n-2)c} (f(n-2)) \\ &= (n-1)! \frac{f}{c}. \end{aligned}$$

Thus  $w_{I_k}$  is in the null space of the Shapley value. It is apparent that the null space has dimension  $n-3$  and the  $w_{I_k}$  are linearly independent. Therefore the  $w_{I_k}$  form a basis of the null space of  $\mathbf{M}$ .  $\square$

This theorem suggests that one may determine the topology of the tree from the null space  $\text{Null}(\mathbf{M})$  of its Shapley transformation  $\mathbf{M}$ . Because every different  $n$ -leaf tree topology divides the leaves differently with respect to at least one leaf (hence producing a different split count), the null space bases will differ in at least one vector. Thus  $\text{Null}(\mathbf{M})$  will distinguish the correct tree topology.

An immediate corollary is that  $\text{Null}(\mathbf{M})$  reveals the location of *cherries*. A pair of leaves  $(i, j)$  is called a *cherry* if they have a common parent. This is the case if and only if the tree spanned by  $i$  and  $j$  does not include an internal edge. Therefore, removing the internal edge that contains the common parent splits  $\mathcal{T}$  into a 2-leaf and an  $(n-2)$ -leaf subtree. Using Theorem 4.4 which determines a specific basis for the nullspace, we may detect which edges include the parent of a cherry. This may be verified in the previous examples.

**Corollary 4.2.** *Let  $\mathcal{T}$  be an unrooted tree with leaves set  $N$  and edge set  $E$ . Let  $w^k := w_{I_k} = (w_1^k, \dots, w_n^k, w_{n+1}^k, \dots, w_{2n-3}^k)$  denote the basis vectors of the nullspace of  $\varphi(v_{\mathcal{T}})$ . Then there is a*



tree  $\mathcal{T}'$  with same leaf set  $\varphi(v_{\mathcal{T}}) = \varphi(v_{\mathcal{T}'})$  in which the pair  $(i, j)$  of leaves form a cherry if and only if there exists  $k'$  such that

$$(10) \quad w_i^{k'} = w_j^{k'} = -\frac{n-3}{2(n-2)}.$$

*Proof.* Inspecting (7) in Theorem 4.4 reveals the equivalence, since in case (and only in case) that deletion of  $I_{k'}$  splits the tree into one with two and one with  $n-2$  leaves, the above stated entries in  $w^{k'}$  prevail.  $\square$

## 5. CHARACTERIZATION OF THE SHAPLEY VALUE OF TREE GAMES

The axioms presented in Section 2 uniquely characterize the Shapley value on the class of all  $n$ -person games. However, the class of  $n$ -person games that are derived from a tree is much smaller. By  $\mathcal{V}^{N,E}$  we denote the class of games arising from some tree with set of leaves  $N$  and edge set  $E$ . For games in  $\mathcal{V}^{N,E}$  we will allow positive as well as non-positive edge weights. Thus,  $\mathcal{V}^{N,E}$  is a linear space and we ask for its dimension.

For a fixed pair  $(N, E)$  define games  $v_k$  ( $k \in E$ ) in the following way:  $v_k$  corresponds to the tree in which edge  $k$  is weighted 1 and all other edges are weighted zero. We call such a game a *basis game*. It is readily checked that the game  $v$  associated with the tree that exhibits edge weights  $\alpha_1, \dots, \alpha_n, \alpha_{I_1}, \dots, \alpha_{I_{n-3}}$  is the linear combination  $v = \sum_{k \in E} \alpha_k v_k$ . Moreover, the family  $(v_k)_{k \in E}$  is linearly independent. Therefore these games form a basis of  $\mathcal{V}^{N,E}$  and  $\dim \mathcal{V}^{N,E} = 2n - 3$ .

Next, we examine a basis game  $v_k$  and ask for a “reasonable” distribution  $\psi(v_k) \in \mathbb{R}^n$ . The total diversity is  $v_k(N) = 1$ . We may interpret zero edge weights on either side of the edge  $k$  as having two groups of species, each one being homogeneous. So a natural property would be that the degree of diversity that we assign to one group does only depend on the fraction of this group (and hence of the fraction of the other group) relative to the whole population. It seems plausible that a group on one side of the edge (relatively) diversifies the population more, the more species there are on the other side of the edge. Thus, we may assume that  $\psi_i(v_k)$  is described by a function that is increasing in the fraction  $f(i, k)/n$ . We formulate these considerations as an additional axiom.

**Axiom (group proportionality on basis games):** For fixed  $N$  and  $E$ , a mapping  $\psi : \mathcal{V}^{N,E}$  is said to satisfy *group proportionality on basis games*, if there is some constant  $d \in \mathbb{R}$  such that  $\psi$  satisfies  $\sum_{i \in \mathcal{C}(i, k)} \psi_i(v_k) = d \frac{f(i, k)}{n}$  for all  $i \in N, k \in E$ .

Thus, with  $\psi$  satisfying this axiom, a group's assigned diversity linearly changes with the other group's fraction of the whole population. Using the new axiom, we get a characterization result on  $\mathcal{V}^{N,E}$ .

**Theorem 5.1.** *For each pair  $(N, E)$  (consisting of leaf set  $N$  and edge set  $E$ ) there is one and only one mapping  $\psi : \mathcal{V} \rightarrow \mathbb{R}^n$  that satisfies Pareto efficiency, symmetry, additivity and group proportionality. This mapping coincides with the Shapley value, i.e.,  $\psi = \varphi$ .*

*Proof.* It is immediately verified that the Shapley value satisfies all the axioms (for group proportionality use 4.1).

Now, let  $(N, E)$  be fixed and  $\psi$  satisfy the axioms. First, we take a basis game  $v_k$  and determine  $\psi$ . By symmetry, we may conclude  $\sum_{i \in \mathcal{C}(i, k)} \psi_i(v_k) = c(i, k) \psi_i(k) = c(j, k) \psi_j(v_k)$  for  $i, j \in \mathcal{C}(i, k) = \mathcal{C}(j, k)$ . Pareto efficiency implies  $v_k(N) = 1 = \sum_{j \in N} \psi_j(v_k) = \sum_{j \in \mathcal{C}(i, k)} \psi_j(v_k) + \sum_{j \in \mathcal{F}(i, k)} \psi_j(v_k) =$

$d\left(\frac{f(i,k)}{f(i,k)+c(i,k)} + \frac{c(i,k)}{f(i,k)+c(i,k)}\right) = d$ . Hence, we obtain  $\psi_i(v_k) = \frac{f(i,k)}{nc(i,k)}$  for any  $i \in N$  and  $k \in E$ . Analogously, we get  $\psi_i(\lambda v_k) = \lambda \psi_i(v_k)$  for  $\lambda \in \mathbb{R}$ . Using additivity and Theorem 4.1,  $\psi$  coincides with the Shapley value on  $\mathcal{V}^{N,E}$ .  $\square$

We close this section with two remarks. First, note that any game arising from a tree with nonnegative edge weights is representable as a linear combination of basis games using nonnegative coefficients. Hence, we may derive a version of Theorem 5.1 for classes of games that actually arise from phylogenetic trees.

Second, Theorem 5.1 provides further justification for the use of the Shapley value to analyze phylogenetic trees. If one wants to distribute the total diversity of a population on its species and the distribution rule should satisfy the above (reasonable) axioms, then the Shapley value is the only possible choice. As symmetry, Pareto efficiency and additivity are rather “obligatory” requirements for a plausible rule, it is the proportionality axiom that provides further insight in the rationale behind the Shapley value. Of course, modification of the group proportionality axiom eventually leads to a different distribution rule based on a different rationale.

## 6. THE CORE OF TREE GAMES

Thus far we have been using the Shapley value to solve tree games. However, another solution concept for  $n$ -player cooperative games that is frequently studied is the *core* of a game, which is the set of all imputations  $\vec{x} \in \mathbb{R}^n$  such that for all coalitions  $S \subseteq N$ ,  $\sum_{i \in S} x_i \geq v(S)$ . In this section we examine the core of phylogenetic tree games.

It is apparent that the core for a single player game is 0 and the core of the two player game is  $\{(x_1, x_2) \in \mathbb{R}^2 | x_1 + x_2 = \alpha, x_1 \geq 0, x_2 \geq 0\}$  so we will derive the core for the three- and four-leaf tree games to gain some intuition about what the core looks like.

**Example 6.1.** *The characteristic function of the three-leaf tree is given in section 3.1. From this we get the following system of inequalities:*

$$\begin{aligned} x_A + x_B + x_C &= \alpha + \beta + \gamma \\ x_A + x_B &\geq \alpha + \beta \\ x_A + x_C &\geq \alpha + \gamma \\ x_B + x_C &\geq \beta + \gamma \end{aligned}$$

*It is apparent that the core consists of the single element  $\vec{\ell}$  which is the leaf weights  $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ .*

So we see that the three-leaf tree has only one element in its core, namely the vector of leaf weights. Now we will look at the four-leaf tree game which will help us see how internal edges affect the core.

**Example 6.2.** *It is easy enough to derive the characteristic function of the four-leaf tree game given in figure 3 so we will not write it here. This game yields the following system of inequalities:*

$$\begin{aligned}
 (11) \quad & x_A + x_B + x_C + x_D = \alpha + \beta + \mu + \gamma + \delta \\
 & x_A + x_C \geq \alpha + \mu + \gamma \\
 (12) \quad & x_B + x_D \geq \beta + \mu + \delta \\
 & \vdots
 \end{aligned}$$

From (11) and (12) we see that

$$\alpha + \mu + \gamma \leq x_A + x_C \leq \alpha + \gamma.$$

So either  $\mu = 0$  in which case we have a degenerate tree and the core is  $\vec{\ell}$  or the core has to be empty since the inequality cannot be satisfied.

From these two examples we obtain the following theorem.

**Theorem 6.1.** *Let  $\mathcal{T}$  be an  $n$ -leaf game tree  $\mathcal{T}$  where  $n \geq 3$ . If the tree is degenerate, then the core consists of the leaf weight vector  $\vec{\ell}$ . Otherwise the core is empty.*

*Proof.* Let  $\mathcal{T}$  be an  $n$ -leaf tree with edge weights  $\alpha_i$  for  $i \in \{1, \dots, 2n-3\}$ . Every tree has at least two cherries, where a *cherry* is a set of two leaves with a common parent. Label the two leaves on one cherry 1 and 2 and label the two leaves on the other cherry 3 and 4 each with corresponding leaf weights  $\alpha_1, \alpha_2, \alpha_3$  and  $\alpha_4$ . We know from the properties of the core that for the set of leaves  $N$ ,

$$(13) \quad \sum_{j \in N} x_j = \sum_{i \in \{1, \dots, 2n-3\}} \alpha_i$$

$$(14) \quad x_1 + x_3 \geq \sum_{k \in P} \alpha_k$$

$$(15) \quad \sum_{j \in N \setminus \{1, 3\}} x_j = \sum_{i \in T \setminus P} \alpha_i$$

where  $P$  is the set of edges in the subtree spanned by A and C. From (13) and (15) we get

$$(16) \quad x_1 + x_3 \leq \alpha_1 + \alpha_3.$$

We know there are no other edge weights included in (16) because the subtree spanned by 2 and 4 (which is included in  $T \setminus P$ ) will have the same edges as  $P$  except for the leaf weights. Thus from (14) and (16) we must have

$$\sum_{k \in P} \alpha_k \leq x_1 + x_3 \leq \alpha_1 + \alpha_3.$$

However this cannot be satisfied and the core is empty unless all of the internal edge weights are zero (i.e., the tree is *degenerate*), in which case the core is the element  $\vec{\ell}$ .

□

Notice that for  $n = 3$ ,  $\mathcal{T}$  is always degenerate, and thus the core will never be empty.

Because the core of tree games is empty in most cases, the Shapley value is a far more interesting solution concept to consider. However, the core has the potential to find (or rule out) degenerate trees easily, unlike the Shapley value.

Suppose we are given the pairwise distances for  $n$  leaves of a tree. If any four leaf subset has an empty core, then the tree is definitely not degenerate. But if any of the inequalities hold then the

subtree spanned by the four leaves in the subset contains a degeneracy. To illustrate this point, see example 6.3.

**Example 6.3.** Consider the 5-leaf tree given in figure 3. Let  $\mu > 0$  and  $\rho = 0$ . Then the four-leaf

subtree ACDE has a nonempty core, namely  $\begin{pmatrix} \alpha + \mu \\ 0 \\ \gamma \\ \delta \\ \epsilon \end{pmatrix}$ . Thus there is a degeneracy among the

leaves ACDE which we can see ( $C, D, E$  all have a common parent). However, in the four leaf subtree ABCE, we have

$$\alpha + \mu + \gamma \leq x_A + x_C \leq \alpha + \gamma$$

so the core is empty. Thus the tree is not totally degenerate but it contains a degenerate subtree CDE.

## 7. CONCLUSION

In this paper we have presented a biological interpretation of the Shapley value on games derived from phylogenetic trees. From a mathematical perspective, we showed how the Shapley value of tree games can be calculated from the Shapley value of the subtrees even if the tree itself cannot be constructed from those subtrees. It is worth noting again the dependence many of our results have on the *split counts*, the division of leaves with respect to a given edge. We have also proved some results about the null space of the Shapley transformation on tree games, as well as the emptiness of the core.

Our work suggests several directions for further research. For instance:

- Can our results be used in some way to assist with reconstruction of trees from data?
- Is there a way to determine split counts from raw data, and can this assist in determining the correct tree topology?
- If there were a way to estimate the Shapley value from data, this would be enough to determine edge weights of a degenerate tree. Do the leaf weights of this tree have any significance?
- If we use the Shapley value to rank the species in the Noah's ark problem for preservation, to what extent can we guarantee that the diversity of the top  $k$  species (i.e., the weight of the subtree spanning them) approximates the total diversity of all  $n$  species? Determine a bound that depends on  $k$  and  $n$ .

## REFERENCES

- [1] William H.E. Day and F.R. McMorris. *Axiomatic Consensus Theory in Group Choice and Biomathematics*. SIAM, Philadelphia, 2003.
- [2] Joseph Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Inc., Massachusetts, 2004.
- [3] Anirban Kar. Axiomatization of the shapley value on minimum cost spanning tree games. *Games and Economic Behavior*, 38:265–277, 2002.
- [4] Klaus Nehring and Clemens Puppe. A theory of diversity. *Econometrica*, 70(3):1155–1198, 2002.
- [5] Lior Pachter and David E Speyer. Reconstructing Trees from Subtree Weights.
- [6] Charles Semple and Mike Steel. *Phylogenetics*. Oxford University Press, New York, 2003.
- [7] Lloyd S. Shapley. A value for  $n$ -person games. In *Ann. Math. Studies*, volume 28, pages 307–317. Princeton University Press, Princeton, N.J., 1953.

- [8] Martin L. Weitzman. On diversity. *Quarterly Journal of Economics*, 107(2):363–405, 1992.
- [9] Martin L. Weitzman. The Noah’s ark problem. *Econometrica*, 66(6):1279–1298, 1998.